



Published in final edited form as:

Science. 2015 May 22; 348(6237): 917–921. doi:10.1126/science.aaa9266.

## Structure of the HIV-1 RNA Packaging Signal

Sarah C. Keane<sup>1</sup>, Xiao Heng<sup>1,2</sup>, Kun Lu<sup>1,3</sup>, Siarhei Kharytonchyk<sup>4</sup>, Venkateswaran Ramakrishnan<sup>1</sup>, Gregory Carter<sup>1</sup>, Shawn Barton<sup>1</sup>, Azra Hosis<sup>1</sup>, Alyssa Florwick<sup>1</sup>, Justin Santos<sup>1</sup>, Nicholas C. Bolden<sup>1</sup>, Sayo McCowin<sup>1</sup>, David A. Case<sup>5</sup>, Bruce Johnson<sup>6</sup>, Marco Salemi<sup>7</sup>, Alice Telesnitsky<sup>4,\*</sup>, and Michael F. Summers<sup>1,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250

<sup>4</sup>Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor MI 48109-0620

<sup>5</sup>Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854

<sup>6</sup>One Moon Scientific, Inc., 839 Grant Ave., Westfield, NJ 07090, and CUNY Advanced Science Research Center, 85 St. Nicholas Terrace, New York, NY 10031

<sup>7</sup>Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610

### Abstract

The 5'-leader of the HIV-1 genome contains conserved elements that direct selective packaging of the unspliced, dimeric viral RNA into assembling particles. Using a <sup>2</sup>H-edited NMR approach, we determined the structure of a 155-nucleotide region of the leader that is independently capable of directing packaging (Core Encapsidation Signal;  $\Psi^{CES}$ ). The RNA adopts an unexpected tandem three-way junction structure, in which residues of the major splice donor and translation initiation sites are sequestered by long-range base pairing, and guanosines essential for both packaging and high-affinity binding to the cognate Gag protein are exposed in helical junctions. The structure reveals how translation is attenuated, Gag binding promoted, and unspliced dimeric genomes selected, by the RNA conformer that directs packaging.

Assembly of HIV-1 particles is initiated by the cytoplasmic trafficking of two copies of the viral genome and a small number of viral Gag proteins to assembly sites on the plasma membrane (1–6). Unspliced, dimeric genomes are efficiently selected for packaging from a cellular milieu that includes a substantial excess of non-viral mRNAs and more than 40 spliced viral mRNAs (7, 8). RNA signals that direct packaging are located primarily within

\*Correspondence: summers@hhmi.umbc.edu; ateles@umich.edu.

<sup>2</sup>Current: Department of Biochemistry, University of Missouri, Columbia, MO 65211.

<sup>3</sup>Current: Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Rd, Tarrytown, NY 10591.

#### SUPPLEMENTARY MATERIALS

Materials and Methods

Tables S1 and S2

Figures S1 to S4

References (33–52)

the 5'-leader of the genome and are recognized by the nucleocapsid (NC) domains of Gag (4). Transcriptional activation, splicing, and translation initiation are also dependent on elements within the 5'-leader, the most conserved region of the genome (9), and there is evidence that these and other activities are temporally modulated by dimerization-dependent exposure of functional signals (6, 10–13).

Understanding the RNA structures and mechanisms that regulate HIV-1 5'-leader function has been based on phylogenetic, biochemical, nucleotide reactivity, and mutagenesis studies (4). The dimeric leader selected for packaging appears to adopt a highly branched secondary structure, in which there are structurally discrete hairpins and helices that promote transcriptional activation (TAR), tRNA primer binding (PBS), packaging ( $\psi$ ), dimer initiation (DIS), splicing (SD), and dimer stability (U5:AUG) (4, 14) (Fig. 1). Although NMR signals diagnostic of TAR, PBS,  $\psi$ , DIS, U5:AUG, and Poly(A) helices have been observed in spectra obtained for the full-length dimeric leader (13, 15) (Fig. 1A), signals diagnostic of a putative SD hairpin have not been detected (colored magenta in Fig. 1A) (15), and there is little agreement among more than 20 different structure predictions for residues adjacent to the helices (4). For example, predictions vary for stretches of residues shown by *in vivo* nucleotide reactivity (16) and crosslinking with immunoprecipitation (CLIP) (17) to reside at or near sites of Gag binding (4). The TAR, Poly(A), and PBS hairpins of the HIV-1 leader are not required for efficient encapsidation (15), and a minimal HIV-1 packaging element, the Core Encapsidation Signal ( $\Psi^{\text{CES}}$ ), exhibits NC binding properties and NMR spectral features similar to those of the intact 5'-leader and is independently capable of directing vector RNAs into virus-like particles (15). To gain insights into the mechanism of HIV-1 genome selection, we determined the structure of  $\Psi^{\text{CES}}$  by NMR.

Contributions of slow molecular rotational motion to NMR relaxation were minimized by substituting the dimer promoting GC-rich loop of the  $\Psi^{\text{CES}}$  DIS hairpin by a GAGA tetraloop (Fig. 1A). This prevented dimerization (Fig. 1B) but did not affect NC binding (Fig. 1C) or NOESY NMR spectral patterns (18), indicating that the modified RNA retains the structure of the native dimer. Non-exchangeable aromatic and ribose  $H_1'$ ,  $H_2'$  and  $H_3'$   $^1\text{H}$  NMR signals were assigned for nucleotides of the U5:AUG, lower-PBS, DIS, and  $\psi$  helices by sequential residue analysis of 2D NOESY spectra obtained for nucleotide-specific  $^2\text{H}$ -labeled samples (18–20) (Fig. 1D). Very long-range A- $H_2$  NOEs ( $^1\text{H}$ - $^1\text{H}$  distances up to  $\sim 7$  Å) were detected in spectra of highly deuterated samples (Fig. 1E) (as observed for proteins (21)), facilitating assignments.

NMR signals that could not be assigned by nucleotide-specific labeling were identified by a fragmentation-based segmental  $^2\text{H}$ -labeling approach we developed, in which differentially labeled 5'- and 3'-fragments of  $\Psi^{\text{CES}}$  were prepared separately and non-covalently annealed (Fig. 2, A and B, and fig S1). The dimer-promoting loop of the DIS hairpin served as the fragmentation site, and was substituted by a short stretch of intermolecular G:C base pairs (Fig. 2A). Differential  $^2\text{H}$  labeling afforded the following fragment-annealed RNAs (fr- $\Psi^{\text{CES}}$ ; denoted 5'-fragment:3'-fragment- $\Psi^{\text{CES}}$ ; D = perdeuterated fragment; superscripts denote sites of protonation, all other sites deuterated; e.g., G = fully protonated guanosines,  $\text{A}^{2r}$  = adenosines protonated at C2 and ribose carbons):  $\text{A}^{2r}:\text{U}^r\text{-}\Psi^{\text{CES}}$ ,  $\text{A}^{2r}\text{C}^r:\text{U}^r\text{-}\Psi^{\text{CES}}$

$G^r:A^{2r}C^r-\Psi^{CES}$ ,  $D:A^{2r}C^r-\Psi^{CES}$ ,  $A:D-\Psi^{CES}$ , and  $D:A-\Psi^{CES}$  (fig. S1). Except for residues at the sites of substitution, the NMR spectra of the fr- $\Psi^{CES}$  RNAs were consistent with those of the parent, non-fragmented RNA. For example, NOEs that correlate A124-H<sub>2</sub> with cytosine and uridine H<sub>1</sub>' protons in 2D NOESY spectra obtained for non-fragmented  $A^{2r}C^r-\Psi^{CES}$ ,  $A^{2r}U^r-\Psi^{CES}$ , and  $A^{2r}C^rU^r-\Psi^{CES}$  samples were also detected in spectra obtained for fragment-annealed  $A^{2r}:U^r-\Psi^{CES}$  and  $A^{2r}C^r:U^r-\Psi^{CES}$  constructs, indicating that A124 resides near a cytosine (C125) in the 5'-fragment and a uridine (U295) in the 3'-fragment (Fig. 2C). More than 80 long-range and sequential A-H<sub>2</sub> NOEs were identified using the <sup>2</sup>H-edited NMR approach (Fig. 2E). <sup>1</sup>H NMR assignments were validated by NOE cross peak pattern redundancy and database chemical shift analyses (18, 22) (fig. S2).

The NMR data indicate that residues proximal to the major splice donor site do not form a hairpin, but instead participate in long-range base pairing within an extended DIS stem and a short helical segment, H1 (Fig. 2E). To determine if this secondary structure is also adopted by the native 5'-leader, we obtained NOESY data for dimeric, <sup>2</sup>H-labeled 5'-leader constructs. Adenosine-H<sub>2</sub> signals diagnostic of the U5:AUG, DIS, PBS, and  $\psi$  helices were observed in spectra obtained for the native leader ( $[5'-L]_2$ ), as expected (15). However, signals diagnostic of H1 were only detectable upon removal of the upper PBS loop (substituted by a GAGA tetraloop;  $[5'-L^{PBS}]_2$ ), which eliminated broad upper PBS signals that overlapped with the A124-H<sub>2</sub> signal of H1 (Fig. 2D). This construct exhibits dimerization, NC binding, and NMR properties similar to those of the intact leader (15), and directs both non-competitive (15) and competitive RNA packaging with near-wild type efficiency ( $94 \pm 4\%$  and  $93 \pm 18\%$ , respectively) (Fig. 2F). Thus, the secondary structure observed for  $\Psi^{CES}$ , including the H1 helix, is also adopted by the 5'-leader.

NOE-restrained structure calculations (18) reveal that  $\Psi^{CES}$  adopts a tandem three-way junction structure (Fig. 3, A-C and fig. S3). The overall shape is quasi-tetrahedral, with the U5:AUG, H1, and  $\psi$ -helices forming a plane that is nearly perpendicular to the plane formed by the H1, PBS, and DIS helices (Fig. 3A). Splice site residues G289 and G290 are base paired with C229 and U228, respectively, and adjacent residues are base paired within or near the H1-PBS-DIS (three-way-2) junction (Fig. 3, B-D), and residues of AUG are base paired within the U5:AUG-H1- $\psi$  (three-way-1) junction (Fig. 3B and D). A227-U291 forms an extended DIS hairpin with two internally stacked but non-paired guanines (G272 and G273) and a G240(syn):G278(anti)-G241(anti) base triple. Sequentially stacked pyrimidines (U230\*U288 and C231\*C287) exhibit broad line widths indicative of millisecond timescale conformational exchange (Fig 3E). These residues appear to function as a flexible hinge that connects the extended DIS hairpin with the tandem three-way junction (Fig. 3D). U307-G330 forms an extended  $\psi$ -hairpin structure that contains three non-canonical base pairs [G310(anti)\*A327(anti), G328\*U309, G329\*U308] and a stacked A-A bulge [A311(anti)-A326(anti)] (Fig. 2E), and a flexible GAGG loop (Fig. 3D). Adenosines A302-A305 exhibit pseudo A-form stacking but are not base paired (Fig. 3B), which supports proposals that genomic adenosine enrichment occurs primarily at non-base paired sites (23). A302 and A303 also make A-minor contacts with the U5:AUG helix (Fig. 3B).

To determine if the tandem three-way junction is evolutionarily conserved, we analyzed published HIV-1 leader sequences that contained full coverage of the 5'-UTR (278 total

sequences). Representatives from B, C, and F1 subtypes were included in the analysis (18). Of the 48 base paired nucleotides at or near the three-way junction, 31 were either strictly (16 sites) or very highly (>99%, 15 sites) conserved, and 13 displayed high (90.2% – 98.9%) identity (Table S2). Only 11 of 126 substitutions resulted in loss of base pairing. The remaining 4 sites – <sup>227</sup>A, <sup>279</sup>G, <sup>286</sup>A and <sup>288</sup>U – exhibited significant variation, ranging from 12% (<sup>288</sup>U) to 50.3% (<sup>227</sup>A). Most changes mapped to terminal branches of the  $\Psi^{\text{CES}}$  phylogeny. Thus the tandem three-way junction structure is highly conserved, and the rare variations that disrupt base pairing are due to transient polymorphisms.

The PBS, DIS, and  $\psi$  helices of  $\Psi^{\text{CES}}$  are consistent with models derived from nucleotide reactivity experiments (4), but the SD structure differs significantly. Recent in-gel chemical probing of resolved monomeric and dimeric leader RNAs (24), and probing studies under solution conditions that favor either the monomeric or dimeric species (25), showed that SD loop residues are relatively unreactive in the dimeric RNA, consistent with the  $\Psi^{\text{CES}}$  structure. Pseudo-free energy calculations indicate that the in-gel reactivity data for the dimeric leader (24) are in better agreement with the  $\Psi^{\text{CES}}$  NMR structure than the proposed model (~25% lower experimental pseudo-free energy (18); fig. S4). These findings support proposals that variations in structure predictions are at least partly due to site-specific structural heterogeneity associated with the monomer-dimer equilibrium (13, 24).

HIV-1 NC binds with high affinity to oligonucleotides that contain exposed guanosines (4, 26, 27).  $\Psi^{\text{CES}}$  contains five unpaired Gs (excluding the non-native GAGA tetraloops), a GGG base triple in the DIS stem, and five additional guanosine mismatches clustered at or near the two three-way junctions (G\*U, G\*A, or G\*G) that could serve as NC binding sites (Fig. 4A). Potential contributions of these “junction guanosines” to NC binding were tested by isothermal titration calorimetric (ITC) studies of G-to-A substituted  $\Psi^{\text{CES}}$  RNAs. Free energy calculations indicate that these substitutions, which include three G\*U to A-U substitutions, should not alter the secondary structure of the RNA (18). Replacement of the  $\psi$ GGAG loop by GAAA eliminated one NC binding site, as expected (27), and substitution of the three-way-1 junction guanosines by adenosines (G116A/G333A/G328A/G329A/G331A) eliminated three additional NC sites (Fig. 4B). Mutation of the unpaired (G226, G292 and G294) and mismatched (G224) three-way-2 junction guanosines to adenosines eliminated one NC binding site (Fig. 4B). The influence of these guanosines on RNA encapsidation was evaluated using a competitive *in situ* RNA packaging assay. Human embryonic kidney 293T cells were co-transfected with plasmids that produce vector RNAs containing the wild type ( $\Psi^+$ , which also encodes for viral proteins) and mutant (Test) leader sequences (18). When co-expressed at similar levels,  $\Psi^+$  and Test vector RNAs with native leader sequences were packaged into HIV-1 virus-like particles with similar efficiencies (Fig. 4C). In contrast, significant packaging defects were observed upon G-to-A mutation of the three-way-2 junction guanosines ( $17\% \pm 2\%$ ), the  $\psi$ -loop and three-way-1 junction guanosines ( $10\% \pm 2\%$ ), or all junction and -loop guanosines ( $5\% \pm 1\%$ ) (Fig. 4C). Our findings indicate that the tandem three-way junction serves as a scaffold for exposing clusters of unpaired or weakly paired junction guanosines, thereby enabling their binding to the zinc knuckle domains of NC.

The  $\Psi^{\text{CES}}$  structure explains biochemical, nucleotide reactivity, and phylogenetic results and suggests a mechanism by which the 5'-leader structure regulates translation and splicing (4). *In vitro* translational activity and chemical reactivity of the AUG residues are suppressed upon dimerization (28), and this can be attributed to sequestration of the 5' end of the *gag* open reading frame within the three-way-1 junction (Fig. 3D). Enhanced *in vitro* translational activity caused by mutations immediately downstream of the major splice donor site (A296/A301U and A293C/U295C/G298) can be explained by destabilization of the H1 helix and, for A296/A301U, stabilization of the SD hairpin (29), both of which should favor the monomer. Mutations in AUG that inhibit genome dimerization and suppress packaging (30, 31) are expected to disrupt base pairing in the U5:AUG helix and  $\psi$ -hairpin stem, thereby destabilizing the tandem three-way junction structure required for Gag binding. *In vitro* splicing activity is also attenuated by dimerization (12, 32), and this can be attributed to sequestration of the major splice site recognition sequence within the three-way-2 junction. Antisense oligonucleotides with complementarity to the SD loop inhibit dimerization, and this is likely due to their ability to competitively block formation of the tandem three-way junction (25).

The  $\Psi^{\text{CES}}$  structure also explains the exquisite selectivity of HIV-1 to package its unspliced genome (1, 2). Residues immediately downstream of the major splice site are base paired within the H1 helix and are thus integral to the formation of the tandem three-way junction structure. Although unspliced and spliced HIV-1 mRNAs contain identical 5' sequences (G1-G289), differences in spliced mRNA sequences derived from 3'-exons would preclude formation of the packaging competent junction structure. Similarly, because SD appears to exist as a hairpin in the monomeric, unspliced 5'-leader (12), it is likely that monomeric genomes are also ignored during virus assembly because they do not adopt the tandem three-way junction structure.

Compared to the proteins of HIV-1, structural information for the viral nucleic acids is sparse. RNAs in general are vastly underrepresented in the structural databanks (99,000 proteins versus 2,700 RNA structures), due in part to NMR technical challenges and difficulties obtaining suitable crystals for X-ray diffraction (19, 20). The fr-RNA  $^2\text{H}$ -edited NMR approach enables efficient segmental labeling without requiring enzymatic ligation. Given the ubiquity of hairpin elements that can serve as fragmentation/annealing sites, this method should be generally applicable to modest-sized RNAs (~ 160 nucleotides).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research was supported by grants from the National Institute of General Medical Sciences (NIGMS, R01 GM42561 to MFS and AT, P50 GM 103297 to MS, BJ and DAC). SB, NCB, and SM were supported by a NIGMS grant for enhancing minority access to research careers (MARC U\*STAR 2T34 GM008663), and SB, JS, NCB, and SM were supported by an HHMI undergraduate education grant. We thank the HHMI staff at UMBC for technical assistance, and Brittany Rife (University of Florida) for advice regarding the phylogenetic analysis. The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: pNL4-3 from Dr. Malcolm Martin. Atomic coordinates have been deposited into the Protein Data Bank with accession code 2N1Q.

NMR chemical shifts and restraints have been deposited into the Biological Magnetic Resonance Bank with accession code 25571.

## REFERENCES AND NOTES

1. Coffin, JM.; Hughes, SH.; Varmus, HE. Cold Spring Harbor Laboratory Press; Plainview, N.Y: 1997.
2. Berkowitz R, Fisher J, Goff SP. *Curr Top Microbiol Immun.* 1996; 214:177.
3. Chen J, et al. *Proc Natl Acad Sci USA.* Aug 11.2009 106:13535. [PubMed: 19628694]
4. Lu K, Heng X, Summers MF. *J Mol Biol.* 2011; 410:609. [PubMed: 21762803]
5. Jouvenet N, Simon SM, Bieniasz PD. *J Mol Biol.* Jul 22.2011 410:501. [PubMed: 21762796]
6. Kuzembayeva M, Dilley K, Sardo L, Hu WS. *Virology.* Apr.2014 454–455:362.
7. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. *J Virol.* Jun.1990 64:2519. [PubMed: 2335812]
8. Nikolaitchik OA, et al. *PLoS Pathog.* Mar.2013 9:e1003249. [PubMed: 23555259]
9. Lever AM. *Adv Pharmacol.* 2007; 55:1. [PubMed: 17586311]
10. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. *Nature Revs Microbiol.* 2004; 2:461. [PubMed: 15152202]
11. Greatorex J. *Retrovirology.* 2004; 1
12. Abbink TEM, Berkhout B. *J Virol.* 2008; 82:3090. [PubMed: 18160437]
13. Lu K, et al. *Science.* 2011; 344:242. [PubMed: 21998393]
14. Abbink TEM, Berkhout B. *J Biol Chem.* 2003; 278:11601. [PubMed: 12458192]
15. Heng X, et al. *J Mol Biol.* 2012; 417:224. [PubMed: 22306406]
16. Wilkinson KA, et al. *PLoS Biology.* 2008; 6:883.
17. Kutluay SB, et al. *Cell.* Nov 20.2014 159:1096. [PubMed: 25416948]
18. Information on materials and methods is available at the Science website.
19. Lu K, Miyazaki Y, Summers MF. *J Biomol NMR.* 2009; 46:113. [PubMed: 19789981]
20. Duss O, Lukavsky PJ, Allain FH. *Adv Exp Med Biol.* 2012; 992:121. [PubMed: 23076582]
21. Koharudin LM, Bonvin AM, Kaptein R, Boelens R. *J Magn Reson.* Aug.2003 163:228. [PubMed: 12914838]
22. Barton S, Heng X, Johnson BA, Summers MF. *J Biomol NMR.* Nov 23.2012
23. van Hemert FJ, van der Kuyl AC, Berkhout B. *RNA Biol.* Feb.2013 10:211. [PubMed: 23235488]
24. Kenyon JC, Prestwood LJ, Le Grice SF, Lever AM. *Nucleic Acids Res.* Oct.2013 41:e174. [PubMed: 23935074]
25. Deforges J, Chamond N, Sargueil B. *Biochimie.* Jul.2012 94:1481. [PubMed: 22365986]
26. South TL, Summers MF. *Protein Sci.* 1993; 2:3. [PubMed: 8443588]
27. De Guzman RN, et al. *Science.* 1998; 279:384. [PubMed: 9430589]
28. Baudin F, et al. *J Mol Biol.* 1993; 229:382. [PubMed: 8429553]
29. Abbink TE, Ooms M, Haasnoot PC, Berkhout B. *Biochemistry.* Jun 28.2005 44:9058. [PubMed: 15966729]
30. Poon DT, Chertova EN, Ott DE. *Virology.* Feb 15.2002 293:368. [PubMed: 11886257]
31. Nikolaitchik O, Rhodes TD, Ott D, Hu WS. *J Virol.* 2006; 80:4691. [PubMed: 16641262]
32. Jablonski JA, Buratti E, Stuani C, Caputi M. *J Virol.* Aug.2008 82:8038. [PubMed: 18550660]

**One Sentence Summary**

An NMR structure of a region of the HIV-1 5'-leader gives insight into how the viral genome is selected for packaging.

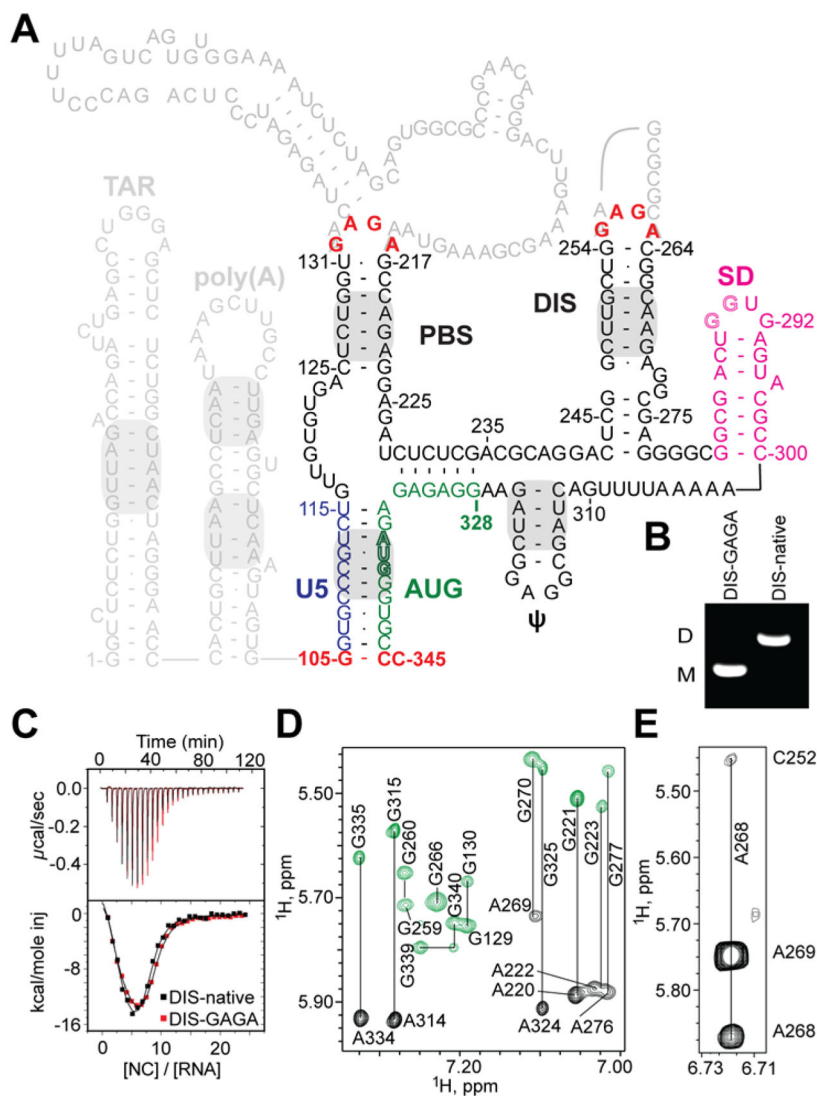
Author Manuscript

Author Manuscript

Author Manuscript

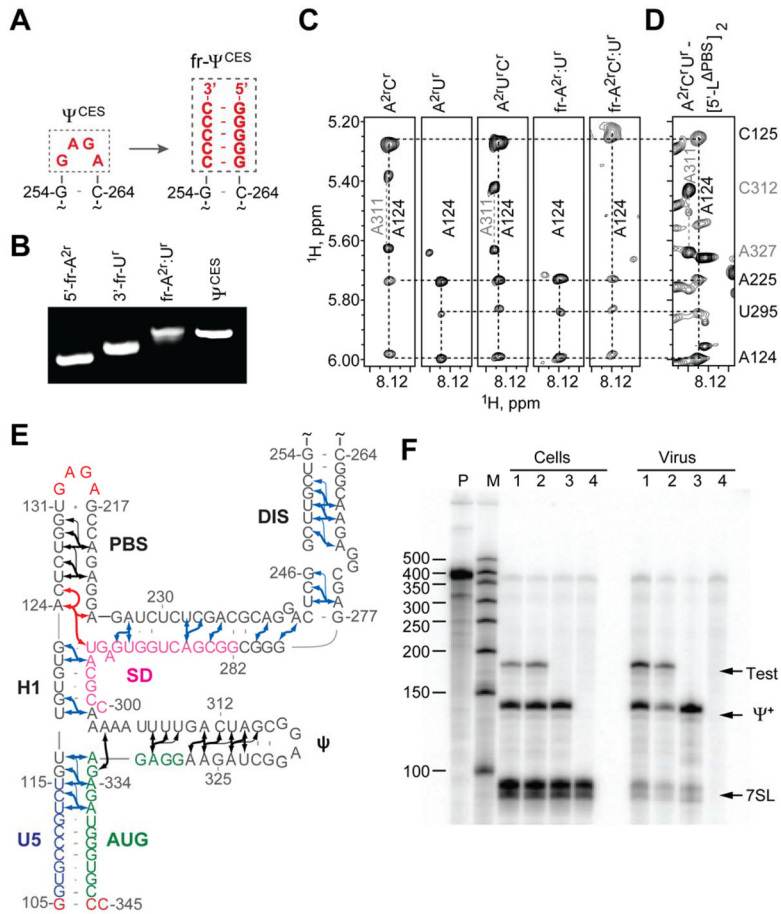
Author Manuscript



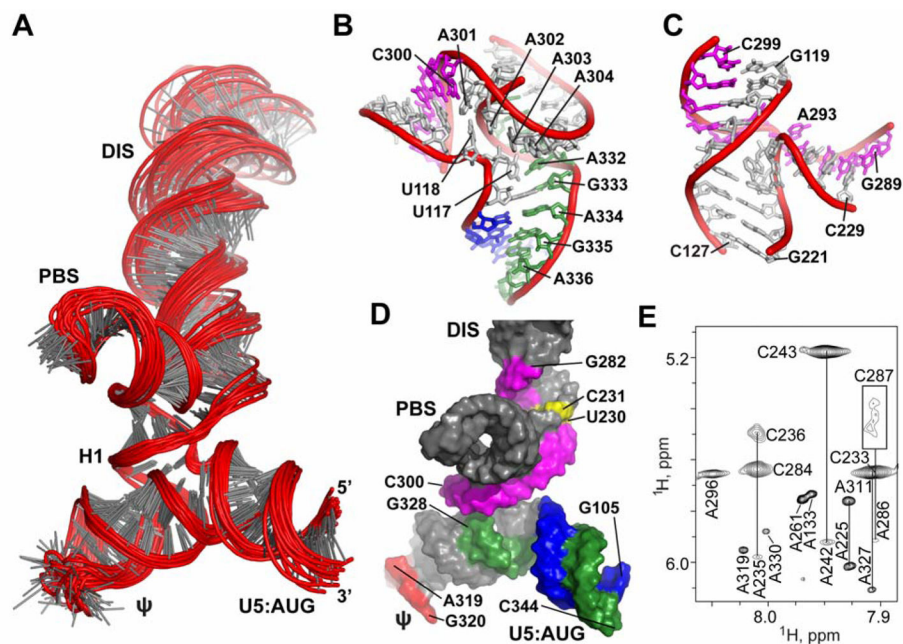


**Figure 1.** HIV-1<sub>NL4-3</sub> 5'-leader and  $\Psi^{\text{CES}}$  RNA construct. **(A)** Predicted secondary structure of the HIV-1 5'-leader (16); gray shading denotes elements detected in the intact leader by NMR (13, 15); dark letters denote  $\Psi^{\text{CES}}$  (non-native residues colored red; see text). **(B and C)** Substitution of the native DIS loop residues (DIS-native) by GAGA (DIS-GAGA) prevents dimerization (B) but does not affect NC binding (C). **(D)** Representative NOESY spectra for  $\text{G}^8\text{A}-\Psi^{\text{CES}}$  (black) and  $\text{G}-\Psi^{\text{CES}}$  (green); lines connect H8 (vertical labels) and  $\text{H}_1'$  (horizontal labels) signals. **(E)** Representative very long-range NOE ( $\text{A}^{2r}\text{C}^r-\Psi^{\text{CES}}$ ,  $\sim 7 \text{ \AA}$  separation) obtained for  $\text{A}^{2r}\text{C}^r-\Psi^{\text{CES}}$ .

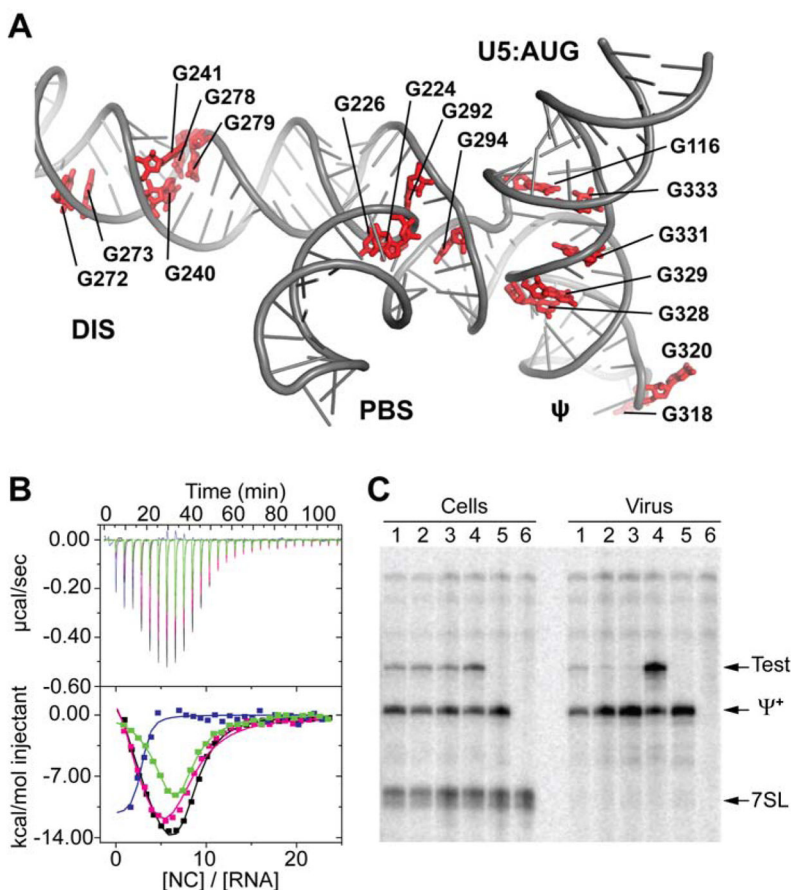




**Figure 2.** Fragmentation-based  $^2\text{H}$ -edited NMR approach and observed  $\Psi^{\text{CES}}$  secondary structure. **(A)** The DIS loop of  $\Psi^{\text{CES}}$  served as the fragmentation site and was substituted by a stretch of intermolecular G-C base pairs. **(B)** Fragment annealing efficiency as measured by native polyacrylamide gel electrophoresis. **(C)** 2D NOESY spectra of uniformly labeled  $\text{A}^{2\text{rC}^{\text{r}}}$ -,  $\text{A}^{2\text{rU}^{\text{r}}}$ -, and  $\text{A}^{2\text{rC}^{\text{r}}\text{U}^{\text{r}}}$ - $\Psi^{\text{CES}}$  and segmentally-labeled  $\text{fr-A}^{2\text{r}}\text{U}^{\text{r}}$ - and  $\text{fr-A}^{2\text{rC}^{\text{r}}\text{U}^{\text{r}}}$ - $\Psi^{\text{CES}}$  samples used to make long-range NOE assignments. **(D)** Similarities in NOESY spectra obtained for  $\text{A}^{2\text{rC}^{\text{r}}\text{U}^{\text{r}}}$ -labeled  $[\text{5}'\text{-L}^{\text{PBS}}]_2$  and  $\Psi^{\text{CES}}$  confirm that the tandem three-way junction structure is present in both constructs. **(E)** NMR-derived secondary structure of  $\Psi^{\text{CES}}$ . Black and blue arrows denote A- $\text{H}_2$  NOEs observable in  $\Psi^{\text{CES}}$  and  $\text{fr-}\Psi^{\text{CES}}$  samples, respectively; red arrows highlight NOEs shown in panel **(C)** and **(D)**; thin arrows denote very long-range NOEs. **(F)** Packaging of native HIV-1 $_{\text{NL4-3}}$  5'-L and 5'-L<sup>PBS</sup> RNAs under competition conditions assayed by means of ribonuclease protection. P, undigested probe; M, RNA sizes marker. Lanes 1 and 2: native HIV-1 $_{\text{NL4-3}}$  helper versus test vectors containing 5'-L<sup>PBS</sup> (1) or native HIV-1 $_{\text{NL4-3}}$  (2). Lane 3: HIV-1 $_{\text{NL4-3}}$  helper expressed without test RNA. Lane 4: mock transfected-cells. Samples obtained from transfected cells (Cells) or viral containing media (Virus) are indicated. Bands corresponding to host 7SL RNA, HIV-1 $_{\text{NL4-3}}$  helper RNA ( $\Psi^+$ ) and co-packaged test RNAs (Test) are labeled.



**Figure 3.** Three-dimensional structure of  $\Psi^{\text{CES}}$ . (A) Ensemble of 20 refined structures (residues 105–344 shown). (B and C) Expanded views of the (B) three-way-1 and (C) three-way-2 junctions. (D) Surface representation of  $\Psi^{\text{CES}}$  highlighting U5 (blue):AUG (green) base pairing and the integral participation of SD residues (pink) in the tandem three-way junction structure. (E) Severe line broadening indicative of slow (millisecond) conformational averaging was observed for stacked, mismatched pyrimidines in the extended DIS stem (yellow in (D); broadened C287- $\text{H}_1'$  signal boxed in (E)). NOE patterns and sharp NMR signals also indicate that the  $\psi$  hairpin loop is unstructured (red in (D)).



**Figure 4.** Junction guanosines mediate NC binding and packaging. (A)  $\Psi^{CES}$  contains 17 unpaired or weakly paired guanosines (red) that serve as potential NC binding sites. (B) Mutation of the three-way-2 (green) or  $\psi$  (magenta) guanosines to adenosines modestly reduces NC binding ( $N = 7.0 \pm 0.3$  and  $7.0 \pm 0.5$ , respectively) relative to WT  $\Psi^{CES}$  (black;  $N = 8.0 \pm 0.3$ ). Mutation of  $\psi$  and three-way-1 guanosines to adenosines (blue) severely inhibits high affinity NC binding ( $N = 2.0 \pm 0.1$ ). (C) Competitive packaging of HIV-1<sub>NL4-3</sub> vectors containing native and mutant 5'-leader sequences, assayed by means of ribonuclease protection. Lanes 1 to 4: native HIV-1<sub>NL4-3</sub> versus test vectors containing 5'-L<sup>3way2-G/A</sup> (1), 5'-L<sup>3way1-G/A</sup> (2), 5'-L<sup>3way1,2-G/A</sup> (3), and 5'-L (4). Lane 5: HIV-1<sub>NL4-3</sub> helper expressed without test RNA. Lane 6: mock transfected-cells. Samples obtained from transfected cells (Cells) or viral containing media (Virus) are indicated. Bands corresponding to host 7SL RNA, HIV-1<sub>NL4-3</sub> helper RNA ( $\Psi^+$ ) and co-packaged test RNAs (Test) are labeled.